基于非对称双分支交互神经网络的水下生物识别*

赵　力，宋　威

(江南大学 物联网工程学院，江苏 无锡 214000)

摘　要：针对水底环境存在着可见度低、光照条件差、物种间特征差异不明显等问题，基于卷积神经网络，提出了一种新的非对称双分支水下生物分类模型。模型中交互分支利用不同的卷积神经网络中间层提取局部特征并通过交互模块对局部特征进行交互，增强分类模型的局部特征学习能力；卷积神经网络分支可以有效地学习到目标的全局特征，弥补交互分支中忽略的全局信息。在 Fish4-Knowledge(F4K)、EILAT、RAMAS 三个数据集上取得了 98.9%、98.3%、97.9%的准确率，较前人方法有显著提高；视觉解释也验证了该模型可以有效地捕捉到局部特征并消除背景影响。最终显示，该模型在水下环境具有良好的分类性能。

关键词：水下生物分类；非对称双分支；交互分支；交互模块；局部特征；卷积神经网络分支；全局特征

中图分类号：TP319　　doi: 10.19734/j.issn.1001-3695.2020.03.0083

Asymmetric two-branch interactive neural network for underwater image classification

Zhao Li, Song Wei

(*School of Internet of Things Engineering*, *Jiangnan University*, *Wuxi Jiangsu* 214000, *China*)

Abstract: Based on convolution neural network, this paper proposed a new asymmetric two branch underwater biological classification model to solve the problems of low visibility, poor illumination conditions and no obvious differences among species in the underwater environment. In the model, the interactive branch used different convolution neural network to extracted local features and interacted with local features through the interactive module to enhanced the classification model. Convolutional neural network branch could effectively learned the global characteristics of the target and made up for the global information ignored in the interactive branch. Finally, this model obtains 98.9%, 98.3% and 97.9% of the accuracy on the three data sets of fish4 knowledge (f4k) , Eilat and RAMAS, which are significantly improved compared with the previous methods. visual interpretation also verifies that the model can effectively capture local features and eliminates the background influence. Finally, it shows that the model has good classification performance in underwater environment.

Key words: subaqueous classification; asymmetric branch; interactive branch; interactive module; local feature; convolutional neural network branch; global feature

0　引言

海洋生物在人类生活中扮演着非常重要的角色，也是人类宝贵的资源之一。经过海洋专家学者几十年的调差研究，我国管辖海域记录到的海洋生物多达 20278 种，其中包括 5 个生物界，44 个生物门，占世界海洋生物总种数的 10%，占总数量的 50%。海洋生物识别用广泛，可用于水产、生物、海洋等环境的研究、开发、管理等。对各类生物进行建立数据库，利用人工智能的方法自动识别生物，不仅有利于海洋生物资源的开发和利用，也能在海洋渔业生产中发挥重要的作用，对学术研究和经济价值都具有重大意义。

利用传统的机器学习进行物种识别过程大致为：获取图像，提取特征，构建分类器，然后将特征输入分类器中进行分类，如：Phenoix 等人[1]采用贝叶斯和高斯核混合模型对鱼类特征进行分层分类的方法来实现分类识别；杜伟东等人[2]提出了一种提取多方位声散射数据的小波包系数奇异值、时域质心及离散余弦变换系数特征；并进行特征融合，最后使用 SVM 进行分类的识别方法；尽管这类方法在基于计算机视觉的海洋生物分类方法研究上取得了重大进展，但是依旧存在明显的不足：分类器性能的好坏很大程度上取决于人为设置的特征是否合理，然而人在选择特征时往往都是依靠经验，具有真大的盲目性和不确定性。

和传统机器学习方法相比，近年来崛起的深度学习方法能够从大量数据中通过卷积等操作自动学习特征，很好地解决了人工选择特征的问题，已经成为解决许多计算机视觉问题的首选；如：Abdelouahid 等人[3]和顾正平等人[4]都提出了采用深度卷积神经网络模型进行鱼类识别的方法,虽然这些方法都在性能上取得了较好的效果，但是依然存在着明显的问题：首先，特征信息在卷积神经网络中传递时存在着信息丢失的现象，而这些模型都注重于对单个卷积层的输出进行分类，因此会丢失一些十分重要的分类信息；其次，在光照不足的水下环境中卷积神经网络容易受到背景的影响。信息丢失和背景影响都会导致分类性能的下降，因此需要在训练时加入大量的额外标注信息，才能取得较好的分类性能，而对数据进行额外标注本身是一项费时且昂贵的工作，所以在实际应用中难以满足，具有很大的局限性。

本文针对上述问题及任务，基于 CNN 提出了一种新的非对称双分支交互神经网络，具有以下结构和特点：a) 交互分支：采用卷积神经网络的中间层提取图像特征，然后通过交互模块对不同中间层所学习到的局部特征进行集成，以增强交互分支对局部特征的捕捉和学习，有效弥补特征信息在传递中出现丢失的不足。b) 卷积神经网络分支：能有效的捕

收稿日期：2020-03-22；**修回日期：**2020-04-25　**基金项目：**国家自然科学基金资助项目(61673193)；中央高校基本科研业务费专项资金资助项目(JUSRP51635B)；中国博士后科学基金资助项目(2017M621625)；江苏省自然科学基金资助项目(BK20181341)

作者简介：赵力，男，河南洛阳人，硕士研究生，主要研究方向为深度学习，计算机视觉(3044242135@qq.com);宋威，江苏无锡人，副教授，硕导，博士，主要研究方向为深度学习，模式识别.

捉到目标的全局信息，弥补交互分支过于注重局部信息而忽略全局信息的不足。c) 两大分支通过融合层相结合，在光照不足的水下环境中也可以良好的捕捉和学习到目标的局部和全局特征信息，并区分目标和背景，消除背景影响；显著提高分类效果。本模型有效的解决了现有传统机器学习模型和现有深度学习模型存在的缺陷，并且在三个数据集上的都有着优于其他模型的分类性能。

1　相关工作

1.1　局部特征学习

特征学习是图像分类过程中十分重要的部分；与基类别(如：猫和狗)之间的差异相比，同种基础类别中的不同种子类别(如：不同种类的珊瑚[18]等)生物体之间的差异非常细微，而且这些细微的差异仅存在于目标图像的局部特征(如珊瑚的冠部、叶片，鱼类的鳍、尾、腹等)；仅仅通过全连接的普通神经层很难解析到这些细微的特征信息，因此，在识别过程中，常规的神经网络模型性能往往会受到限制[5]。针对上述问题，Zhang 等[32] 提出了一种能够从卷积特征中挑选出具有分辨力的局部特征的算法,利用 Selective search 产生候选局部区域，然后利用 MMP (Multi-max pooling) 方法，直接从候选的局部区域中产生局部特征，对这些特征做聚类，并计算每一个聚类簇的重要性，选择重要的聚类簇作为最终的图像局部特征表示；Perronnin 等人[33]利用 FV (Fisher vector) 编码将目标图像的所有候选局部特征表示成一个向量，使用高斯混合模型 (Gaussian mixture model, GMM) 对候选局部特征进行聚类,并通过计算各个类的相互信息值选取重要的局部特征使网络进行学习；Simon 等[34]利用卷积网络特征产生的关键点，并基于这些关键点来提取局部特征信息。以上方法虽然能有效的提取到局部特征，但均采用 Selective search 方法产生候选局部区域，并需要计算各聚类簇之间的重要性，因此面临巨大的计算代价问题。

Lin 等[6]提出了一种对两个独立 CNN 的输出特征进行融合的方法，将两个 cnn 的输出特征向量进行外积然后产生高维特征，进入全连接层进行分类；Kong 等人[7]在此基础上对方差矩阵采取低秩化，降低了计算复杂度；Maji 等[8]提出了矩阵平方根归一化，进一步提升了在分类上的性能；Wei 等[9]认为常用的 1*1 卷积核对特征进行降维会导致降维后的特征多样性降低，所以采用 P 奇异向量降维；Gao 等[29]利用 Tensor Sketch 对二阶信息进行统一并减小特征维度；Cui 等[20]在此基础上使用 Tensor Sketch 将高阶信息进行汇总；Gou 等[10]通过对特征矩阵增广的方法得到了同时包含一阶和二阶信息的特征，并利用 tensor sketch 对其进行融合操作；但是，这些方法仅仅考虑对来自单个卷积层的输出特征进行处理，而在实验中发现在 CNN 中不同的卷积层学习到的特征并不相同，且这些特征信息在通过不同卷积层时会发生明显的信息丢失，因此单个卷积层的输出特征图并不能很好地表现局部特征之间的细微差异。

本文方法将利用多个卷积层提取图像特征，并通过交互的形式集成各个卷积层捕捉到的局部信息，以此来增强模型对细微局部特征的捕捉和学习能力。

1.2　卷积神经网络

由于近年来深度学习在各领域取得良好成果，CNN(卷积神经网络)已经成为各种视觉识别任务的通用特征提取器；Chatfield 等人[11]以 VGGnet 基于图像分类对 CNN 的性能进行了评估，并和以前的特征编码方法进行了比对；实验表明，更深的 CNN 表现优于在已增强数据上训练的深度较浅的 CNN 模型。

作为以水下为应用场景的分类任务，尽管水下图像存在着光照、颜色、角度等诸多因素的影响，但 CNN 依旧证明了它在图像分类领域的优势[12~14]；然而这些方法也存在着明显的缺陷：CNN 在对目标的特征提取时，往往容易受到背景影响，误将背景噪声作为目标进行信息提取，因此在训练时，需要加入人工制作的诸如形状、颜色和纹理等手工特征信息，以加强模型对目标和背景的区分能力；对手工特征信息的依赖导致这些方法很难应用于大型数据集,具有一定的局限性；而本文提出的分类方法，不需要依靠任何人工特征信息便可以有效的除去背景影响，可以应用于任何水下图像数据集。

2　非对称双分支网络

在本章中，基于卷积神经网络，提出了一个非对称的双分支网络用于水下物种分类，不需要依赖任何人工特征信息就可以消除背景产生的影响并捕捉到细微的局部特征；适用于何种水下场景的生物识别数据。

2.1　非对称双分支

与普通的基类别识别不同，同一基类别的不同子类别之间通常具有相似的外观，各个类别间的差异更加细微，子类别识别只能通过微小的局部特征差异进行区分，因此，如何提取并有效学习到局部特征信息, 成为了决定子类别识别算法成功与否的关键所在。但是绝大多数卷积神经网络模型仅仅专注于利用单卷积层进行特征学习，而完全忽略了特征信息在不同层之间传递时发生的信息损失；所以每个卷积层学习到的特征信息是不完整的，因此为了能捕捉到更多的局部特征，本方法在卷积神经网络的基础上加入了一种交互非分支；如图 1 所示，本模型由交互分支和 CNN 分支构成；其中交互分支使用 3 个以卷积层为主的特征提取器，对图像进行特征提取，然后将不同提取器提取的特征输入交互模块，以增进不同特征之间的信息交互；与基类别识别相比，子类别识别更加注重对局部特征的学习，图像信息的信噪比更低，因此更容易受到光照、姿态、背景等因素的影响；而 CNN 分支可以有效的对目标图像的全局信息(如目标的形状、外观等)进行提取，增强模型对图像中目标的定位能力，消除光照、背景等因素的影响，以弥补交互分支注重局部而忽略全局信息的不足；两大分支的输出最终通过融合层加权进行集成。

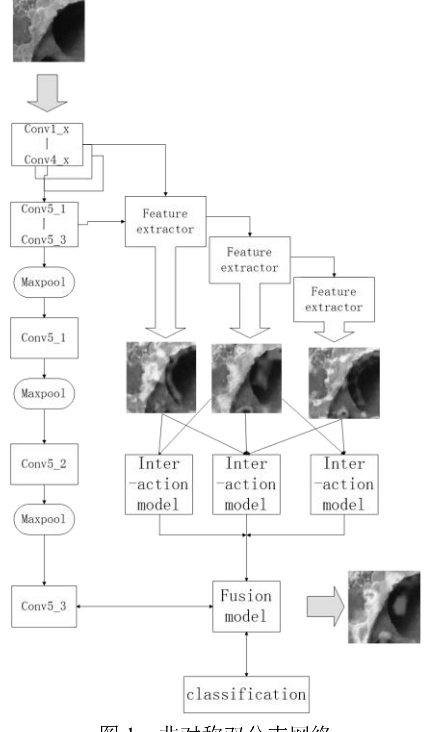


图 1　非对称双分支网络

Fig. 1　Asymmetric two-branch network

2.2　交互分支

交互分支主要由特征提取器、交互模块构成，本分支目的在于捕捉目标图像中细微的局部特征。在分类中这些细小的局部特征具有很强的代表性，因此可以有效提高模型的分类性能。

2.2.1 交互分支分解

Kim 等人 [16]提出了使用 Hadamard 乘积的因式分解，用于多模式学习的有效注意力机制。在本小节，简要介绍因式分解的基本公式；假设一个图像 I 通过以卷积层为主的特征提取器进行过滤，提取器的输出为高度 H ，W 宽度，C 通道的特征映射 $X \in R^{H \times W \times C}$;将 X 中空间上的 c 维描述符表示为 $x=[x_1,x_2,\ldots,x_c]^T$。那么交互模型可以被定义为

$$z_i = x^T W_i x \tag{1}$$

其中 $W_i \in R^{c \times c}$ 是权重矩阵，z_i 是模型输出。根据 Rendel[17]提出的矩阵分解，式(1)可以分解为两个 1 维向量：

$$z_i = x^T W_i x = x^T U_i V_i^T x = U_i^T x \circ V_i^T x \tag{2}$$

其中 $U_i \in R^c, V_i \in R^c$ 。假设 o 维的交互分支输出 Z=[z1,z2,…,zo]，则 $z \in R^o$ 定义为

$$z = U^T x \circ V^T x \tag{3}$$

其中 $U \in R^{c \times d}, V \in R^{c \times d}$ 是不同交互模块的权重矩阵，。为 Hadamard 积，d 为决定交互层性能和计算复杂度的可定义尺寸参数。

2.2.2 交互模块

交互模块目的在于增进不同特征提取器所提取到的特征图之间的交互性：首先通过独立的非线性映射将来自不同提取器的特征扩展到高维空间，以便于卷积层捕捉不同目标局部的特征，然后通过 hadamard 积对逐元素进行集成，以达到不同的局部特征之间进行交互的目的；最后，执行求和，将高维特征压缩为紧凑特征。

单个交互模块中，对空间上第 i 维的不同特征使用 hadamard 积进行交互，可以定义为

$$z_i = U_i^T x \circ V_i^T y \tag{4}$$

其中 x，y 为来自于不同提取器所提取的特征，U_i^T 和 V_i^T 为映射矩阵。最后对整个空间上的特征矩阵执行求和，将高维特征压缩为紧凑的特征向量，假设空间维度为 o，则写作：

$$Z = z_1 + z_2 + z_3 + \ldots + z_o = U^T x \circ V^T y \tag{5}$$

在分支中加入了多个交互模块以集成多个特征，从而进一步增强特征信息在分类中的表达能力；假设 x, y, z 分别来自于不同提取器的特征，对于加入多个交互模块的分支中，交互分支输出，即提取到的局部特征 $Z_{interact}$ 为

$$\begin{aligned} Z_{interact} &= Interaction(x,y,z) = \\ &concat(U^T x \circ V^T y, U^T x \circ W^T z, W^T z \circ V^T y) = \\ &U^T x \circ V^T y + U^T x \circ W^T z + W^T z \circ V^T y \end{aligned} \tag{6}$$

其中 $U, V, W \in R^{c \times d}$ ，d 为交互模块中神经层的尺寸参数，决定着交互模块的性能。

2.2.3 特征提取器

如图 2，因为卷积层本身具有提取特征信息的功能，所以将卷积神经网络的中间卷积层取出，加入非线性函数(Relu)和归一化(batch normalization)，作为本模型的特征提取器。

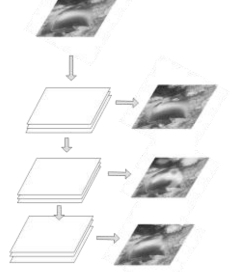


图 2　特征提取器

Fig. 2　Feature extractor

2.3　卷积神经网络分支

上节中的交互分支主要注重对局部特征的学习，所以容易忽略全局信息(例如目标的形状，外观等)，导致模型对目标的定位能力较弱，在识别过程中容易受到背景和光照等因素的影响，因此全局信息在子类别识别中也有着至关重要的作用。普通的卷积神经网络对于细微的局部特征提取和学习能力虽然较弱，但可以有效的捕捉到目标图像的全局信息，因此在另一分支中保留了完成的卷积层和池化层，用于提取全局信息以弥补交互分支忽略全局信息的不足；并在融合层中对两个分支的输出赋以不同的权值进行整合：

$$Z_{output} = P^T fusion(Z_{interact}, Z_{object}) = P^T(w_1 \times Z_{interact} + w_2 \times Z_{object}) \tag{7}$$

其中: Z_{object} 为卷积神经网络提取到的全局信息；w_1 、w_2 为局部特征信息 $Z_{interact}$ 与 Z_{object} 对应的权值，总和为 1，以控制局部特征信息和全局信息在分类信息中所占的比重。

实验　在实验中，首先，采用三个最常用的数据集用来评估模型性能，并提供了与前人方法的比较；然后对本模型的各个部分单独进行了评估，最后用视觉解释直观的对模型作出解释。

3　实验

3.1　数据集

采用三个水下生物领域最常用的数据集；

EILAT 数据集[18]：该数据集为从同一相机拍摄的全尺寸图像中提取的图像块，包含 1123 张图像，均为红海 EILAT 岛附近的珊瑚礁调查中相机捕捉到的 64*64 像素全尺寸图片；并由专家标记分为 8 类。本数据集采用 10 折交叉验证方法，即其中 90%的图像构成训练集，其余 10%作为测试集。

Rosenstiel 海洋与大气科学学院在珊瑚礁调查中收集的，包含 766 个图像，被专家标记为 14 个类别，每个图像尺寸为 256*256.该数据集与 EILAT 数据集使 RAMAS 数据集[18]:该数据集是迈阿密大学用相同的交叉验证方法。

Fish4-Knowledge (F4K)[19]数据集：该数据集是台湾电力公司、台湾海洋研究所和垦丁国家公园在 2010 年 10 月 1 日至 2013 年 9 月 30 日期间，在台湾南湾兰屿和胡比胡的水下观景台收集的影像数据，其中包含 23 种鱼类，大小为 20*20 至 200*200，共 27370 张鱼类的水下图像。该数据集的 80%构成训练集，其余 20%作为测试集。

3.2　实验设定

选用在 ImageNet 分类数据集上预训练的 VGG-16 作为模型中的卷积神经网络分支(非对称双分支模型亦可使用其他种类的卷积神经网络，例如 Inception,Resnet 等)，除去其最后三个全连接层，然后加入本文提出的交互分支；输入图像尺寸统一为 224*224，为了证明本模型的性能并非依赖于高超的预处理手段，仅采用最简单的数据增强方法，如随机水平翻转，随机平移；在训练过程中使用 batch size 为 16 的随机梯度下降(SGD)方法调整整个模型，并将 momentum(动量)设定为 0.9，weight decay (权重衰减)为 $5*10^{-3}$，learning rate(学习率)为 10^{-3} 并在学习停滞时减少 10 倍；融合层中调节局部特征信息和全局信息所在比重的权值 w_1 和 w_2 初始值均设为 0.5 以保证二者占比均衡，随后在实验中加以调整；所有实验均在谷歌深度学习框架 tensorflow 上进行实现。

3.3　中间层选取

作为特征提取器的 CNN 中间层决定了捕捉到的局部特征在分类中是否具有代表性；所以中间层的选取十分重要。以 VGG16,VGG19 为例，将每个池化层之前的神经层视为一个完整的卷积模块(每个卷积模块包含 2-3 个卷积层)，并将每个卷积模块的输出分别输入 softmax 层进行了分类，在三个数据集上的精度如图 3 所示；可以看出在 CNN 中，随着

层数的加深，卷积层所提取到的特征级别也越高，在分类中也更具有代表性；所以在 CNN 中，深层提取到的特征比浅层所提取到的特征更加有利于分类，这也与前人研究所得结论一致[20];所以选取最后一个卷积模块中的中间层作为特征提取器。

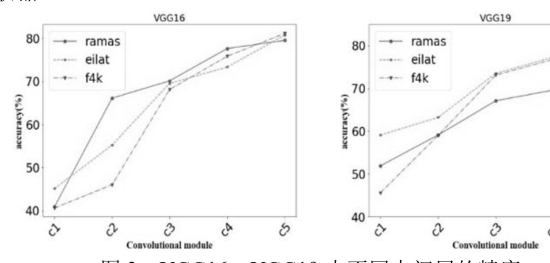


图 3　VGG16、VGG19 中不同中间层的精度

Fig. 3　Accuracy of different intermediate layers in VGG16、VGG19

3.4　交互模块的映射尺寸参数及定量分析

式(6)中提到，交互模块中存在着决定交互性能的神经层尺寸参数 d。为了选取合适的 d，用不同的模块组合在 EILAT 数据集上进行了实验，结果如图 4 所示；其中 I(1,2)+CNN 表示将提取器 1,2 输出交互层然后将其输出和左分支的 CNN 输出进行融合并进行分类；I(1,3)、I(2,3)代表同上；F+CNN 代表 I(1,2)+ I(1,3)+ I(2,3)+CNN;

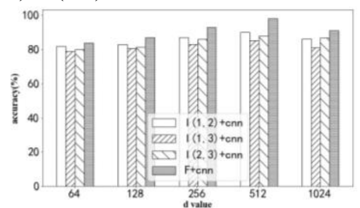


图 4　各种组合在不同尺寸 d 时的性能

Fig. 4　Performance of various combinations at different sizes d

图 4 中可以看出：首先，在无论尺寸 d 取何值，F+CNN 的组合性能均优于其他组合，这表明可以通过多个交互模块以增强分类性能；其次，当尺寸 d 从 64 增加到 512 时，所有的组合性能均有提升，但增加至 1024 时，性能有所下降。考虑到：过大的 d 值会产生更高的计算复杂度，d 过小又会导致模块性能下降。因此，在接下来的实验中均选取 d=512 为最佳尺寸。

为了对交互模块进行定量，分别选用了包含 1-3 个交互模块的组合 RAMAS、EILAT 两个数据集上进行了性能评估；结果如表 1 所示。

表 1 在不同数据集上的各数量模块的组合性能评估

Tab. 1　Combined performance evaluation of each number of modules on different datasets

组合方法	RAMAS	EILAT
I(1,2)+CNN	84.1	90.5
I(1,3)+CNN	86.3	85.9
I(2,3)+CNN	83.5	87.2
I(1,2)+I(1,3)+CNN	90.9	93.0
I(1,2)+I(2,3)+CNN	91.5	94.5
I(1,3)+I(2,3)+CNN	93.1	95.4
I(1,2)+I(1,3)+I(2,3)	95.5	96.1
I(1,2)+I(1,3)+I(2,3)+CNN	97.9	98.3

首先，综合前 6 项和最后一项可以看出，提高交互模块数量，能明显改善分类性能；其次，最后两项中，交互分支在与 CNN 分支的输出融合后，性能有所提升，也证明了独立的交互分支过于注重局部信息而忽略了全局信息，而 CNN 分支提供的全局信息在分类中也起到了重要的作用。

3.5　中间层性能分析

为了证明局部特征信息的交互和集成能提高分类性能，在所有数据集均使用相同标准的分割方法情况下，将三个特征提取器(CNN 中取出的 3 个中间层)的输出，卷积神经网络分支的输出都分别输入全连接层进行了分类测试，并与完整模型的分类(融合层)进行了比较。

为了验证本模型可以使用不同类别 CNN 的想法，在实验中使用了 2 中全完不同的 CNN，比较结果如表 2 所示，值的注意的是，第三个提取器在性能上明显劣于前 2 个提取器，从此可以推断出：虽然更深的卷积层更能提取到有利于分类的抽象特征，但随着深度增加，层之间的信息传递中存在着明显的特征信息丢失，从而导致分类性能下降;此推断也在后文中的视觉解释实验中得到了证实。因此想要提高分类性能，对多个特征信息进行更好地学习才是关键。

表 2　CNN 分别选用 VGG16、resnet 的结果比较

Tab. 2　Comparison of the results of CNN using VGG16 and ResNet

数据集	CNN	特征提取器 1	特征提取器 2	特征提取器 3	融合层
EILAT	80.7	80.1	82.6	81.2	98.3
RAMAS	79.5	82.7	81.4	79.9	97.9
F4K	81.1	83.9	84.5	81.9	97.1
EILAT	81.2	83.3	80.1	80.0	97.2
RAMAS	80.1	81.7	82.6	80.9	96.8
F4K	81.9	82.3	80.0	81.6	96.7

3.6　融合层的权值分析

式(7)中提到，融合层中的权值 w_1 和 w_2 分别控制着局部信息和全局信息在最终分类信息中所占的比重，二者之和恒为 1.为了选取合适的 w_1 和 w_2，以分类精度作为衡量标准，使用不同的 w_1 值在三个数据集上进行了实验，实验结果如图 5 所示。

图 5 中可以看出：首先，当局部信息权值 w_1 取值在 0~0.7 之间时，分类精度随着 w_1 的增加而提升，这表明局部信息在分类过程中至关重要，加入局部信息可显著提升模型的分类性能；其次，当 w_1 取值在 0.7~1 之间时，分类精度随着全局信息权值 w_2 的减小(w_2=1-w_1)而降低，这表明忽略全局信息会使模型在分类过程中受到背景和光照等因素的影响，导致分类性能下降，因此加入适当比例的全局信息可消除这些影响，进一步提升模型的分类性能。考虑到：过小的 w_1 权值会降低模型对局部信息的学习能力，而过大的 w_1 权值又会导致模型对全局信息的忽略。因此，选取 0.7 和 0.3 作为权值 w_1 和 w_2 的最佳取值。

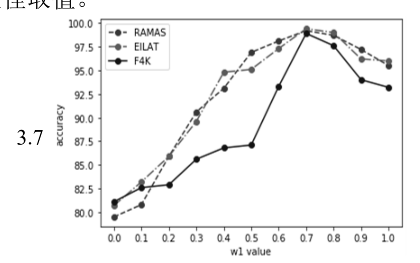


图 5　不同权值(w_1)时模型的分类性能

Fig. 5　The classification performance of the model with different weights (w_1)

3.7　分类结果对比

在本节中，在每个数据集上，都将交互双分支模型的性能与前人方法进行比较；需要说明的是，不同方法及模型的学习能力不同，因此所学习到的视觉特征也并不相同，而模型是否能达到良好的分类性能取决于该模型能否学习到关键的视觉特征。表 3 中显示了双分支交互网络的结果和目前在 RAMAS、EILAT 两个数据集上精度最高的结果；其中 VGG16 和 Resnet-50 均为在 Imagenet 进行预训练然后在数据集上进行训练的结果；值得一提的是第 8 项所采用的方法在训练过程中也依赖手工制作的特征；第 9 项为目前的最新技术，达

到了目前的最高精度，但该方法仅注重局部特征，而忽略了目标图像的全局特征，因此在光照条件差的水下环境中，分类性能受到了背景的影响；表 3 最后一项可以看出，不依赖任何手工特征的交互双分支网络的分类性能明显优于其他方法，达到了最高的分类精度。

表 3　在 RAMAS，EILAT 数据集上的各个方法性能评估

Tab. 3　Performance evaluation of each method on RAMAS, EILAT dataset

方法	RAMAS	EILAT	方法	RAMAS	EILAT
文献[22]	69.3	87.9	文献[21]	85.4	69.1
文献[24]	73.9	67.3	文献[18]	96.5	96.9
VGG-16	79.5	80.7	文献[25]	97.1	97.5
Resnet-50	80.1	81.2	cResFeats[30]	98.8	99.1
文献[23]	82.5	75.2	交互双分支网络	99.2	99.4

表 4 中显示了双分支交互网络的结果和目前在 F4K 数据集上精度最高的结果；其中 VGG16 和 Resnet-50 均为在 Imagenet 进行预训练然后在数据集上进行训练的结果；其中，Wei 等人[27]以 F4K 中的鱼类名称为关键字，使用谷歌搜索引擎下载了更加清晰的图片，并对错误的图像以及边界区域进行了手工删除和切割，以此来构建了高质量的数据集，并在高质量的数据集中达到了 97.3%的精度，而在原有的 F4K 数据集上精度仅有 17.14%；张俊龙等[31]也对每种鱼类的图片按照清晰度和背景影响程度进行了划分，将数据集分为高、中、低品质的三个子数据集，并在高质量数据集上达到了 97.0%的精度，而在中、低质量数据集上的精度为 94%和 90%；顾正平等人[4]采用迁移学习的 CNN+SVM 方法，达到了 98.6%的精度，为本数据集上目前所达到的最高精度；对比前人的各项方法，交互双分支网络的准确度均有明显提高，并且不依赖任何人工提取和制作的特征。

表 4　在 F4K 数据集上的各个方法性能评估

Tab. 4　Performance evaluation of various methods on the F4K dataset　　/%

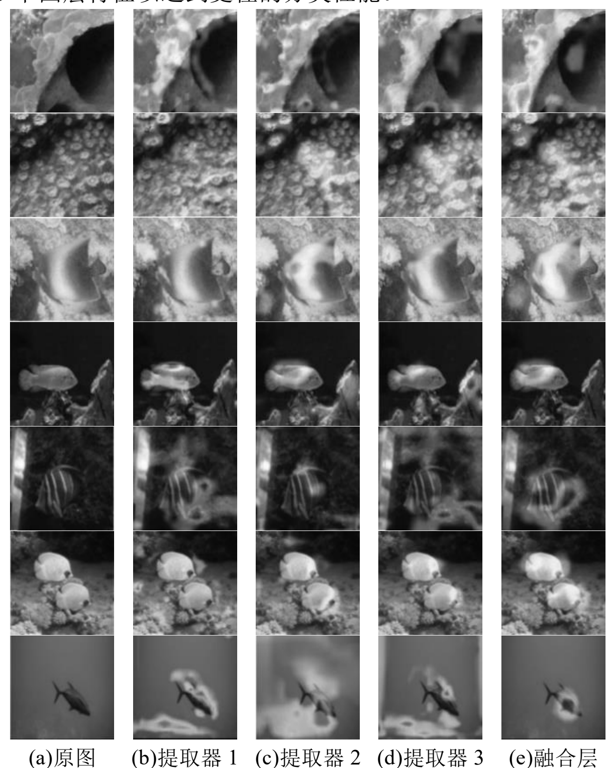
方法	精度	方法	精度
LDA+SVM[26]	80.4	文献[27](高质量数据集)	97.3
VLFeat Dense-SIFT[26]	93.5	文献[31]	97.0
VGG-16	81.1	文献[4]	98.6
Resnet	81.9	交互双分支网络	98.9
文献[15]	96.3		

3.8　视觉解释及分析

为了更好的解释本模型，使用 Grad-cam[28]算法对不同层的输出进行了视觉解释，如图 6 所示，热力部分代表着分类中模型感兴趣的区域，即模型的注意力区域；热力颜色越深的区域，则说明在分类中的贡献越大，所占比重就越大。图 6 表明，在光照条件较差的水下场景，普通的卷积层在提取特征时，由于色差较弱的原因，很容易收到目标背景的影响，误把背景当做目标进行特征提取。此外，还可以看出，在特征提取和学习时，卷积层对局部特征具有一定的捕捉能力，但不同的卷积层所关注区域有很大不同，并且在特征信息的传递中，存在着明显的特征信息丢失；而这些丢失的局部特征(如鱼类的头部、尾部、腹部，珊瑚的冠部、叶片等)在分类中确发挥着至关重要的作用；同时也证明了 3.5 中特征信息丢失而导致分类性能下降的推断。本模型通过促进不同的特征交互并对其进行集成，极大的提高了特征信息的利用率，并有效的去除了由于水下光线弱而导致的背景影响。

4　结束语

本文基于深度学习研究了水下生物分类的算法，并基于卷积神经网络提出了具有交互和集成模块的非对称交互双分支网络模型。通过实验，在三个最常用的水下生物数据集上达到了最高的精度，这充分说明了本模型并不依赖任何人工方法，更不需要有关海洋生物领域的相关知识，便可以达到良好的分类性能。将来，将继续扩展研究，在环境条件更加恶劣、图像质量更差的场景中，如何更加有效的学习并集成多个图层特征以达到更佳的分类性能。



(a)原图　(b)提取器 1　(c)提取器 2　(d)提取器 3　(e)融合层

图 6　不同层输出的视觉解释

Fig. 6　Visual interpretation of different layer outputs

参考文献：

[1] Huang P X, Bastiaan B, Fisher R. Hierarchical classification with reject option for live fish recognition. Machine Vision and Applications 2015, 26 (1): 89-102.

[2] 杜伟东，李海森，魏玉阔. 基于SVM 的多方位声散射数据协作融合鱼分类与识别 [J]. 农业机械学报, 2015, 61 (3): 39-43. (Du Weidong, Li Haifeng, Wei Yukuo. Multi-azimuth acoutic scattering data cooperative fusion using SVM for fishclassification and identification [J]. Transactions of The Chinese Society of Agricultural Machinery, 2015, 61 (3): 39-43.)

[3] Tamou A B, Benzinou A, Nasreddine K, *et al.* Underwater Live Fish Recognition by Deep Learning [C]. International Conference on Image and Signal Processing. Springer, Cham, 2018, 171 (6): 275-283.

[4] 顾郑平，朱敏. 基于深度学习的鱼类分类算法研究 [J]. 计算机应用与软件, 2018, 35 (1): 200-205. (Gu Zhengping, Zhu Min. Fish classifion algorithm bassed on depth learning [J]. Computer Application and Software, 2018, 35 (1): 200-205)

[5] Yandex A B, Lempitsky V. Aggregating local deep features for image retrieval [J]. IEEE International Conference on Computer Vision, Computer Science. 2015: 1269–1277.

[6] Lin Tsungyun, Roychowdhury A, Maji S. Bilinear CNN models for fine_grained visual [C]// IEEE International International Conference on Computer Vision, 2015: 1449-1457.

[7] Kong shu, Charless F. Low-rank bilinear pooling for fine-grained classification [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7025–7034.

[8] Lin Tsungyun, Maji S. Improved bilinear pooling with CNNs [C]// The

British Machine Vision Conference, 2017.

[9] Wei Xing, Zhang Yue, Gong Yihong, *et al.* Grassmann pooling as compact homogeneous bilinear pooling for fine-grained visual classification [C]// European Conference on Computer Vision, Computer Vision. 2018: 365-380.

[10] Gou Mengran, Xiong Fei, Octavia I C, *et al.* MoNet: Moments embedding network [C]// Conference on Computer Vision and Pattern Recognition, 2018: 3175-3183.

[11] Ken C, Karen S, Andrea V, *et al.* Return of the devil in the details: Delving deep into convolutional nets [C]. /The British Machine Vision Conference, Computer Science, 2014.

[12] Khan H, Munawar H, Mohammed B, *et al.* Cost Sensitive Learning of Deep Feature Representations from Imbalanced Data. IEEE Transactions on Neural Networks and Learning Systems, 2017.

[13] Mahmood A, Bennamoun M, An Senjian, *et al.* Deep image representations for coral image classification [J]. IEEE Journal of Oceanic Engineering, 2018: 121-131

[14] Uzair N, Mohammed B, Ferdous S, *et al.* Deep Fusion Net for Coral Classification in Fluorescence and Reflectance Images [J]. Digital Image Computing: Techniques and Applications, 2019.

[15] Rathi D, Indu S, Jain S, *et al.* Underwater fish species classification using convolutional neural network and deep learning [J]. International Conference of Advances in Pattern Recognition, 2017.

[16] Kim J H, On K W, Kim J, *et al.* Hadamard product for low-rank bilinear pooling [J]. International Conference on Learning Representations. 2017

[17] Rendle S. Factorization machines [J]. Interational Conference on Data Ming, 2010: 559-1000.

[18] Shihavuddin A, Gracias N, Garcia R, *et al.* Image-based coral reef classification and thematic mapping [J]. Remote Sens, 2013, 5: 1809-1841.

[19] Boom B J, Huang P X, He Jiyin, *et al.* Supporting ground-truth annotation of image datasets using clustering [C]// International Conference on Pattern Recognition. IEEE, 2012: 1542-1545.

[20] Cui Yin, Zhou Feng, Wang Jiang, *et al.* Kernel Pooling for Convolutional Neural Networks [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. [S. I.] : IEEE, 2017.

[21] Oscar B, Peter J E, David I K, *et al.* Automated annotation of coral reef survey images [C]// IEEE Conference on Computer Vision and Pattern Recognition, [S. I.] : IEEE, 2012: 1170-1177.

[22] Marcos S, Saloma C A, Soriano M, *et al.* Classification of coral reef images from underwater video using neural networks [J]. Opt Express, 2005, 13 (22): 8766-8771.

[23] Stokes M D, Deane G B. Automated prosessing of coral reef benthic images [J]. Limnol Oceanogr Meth. 2009. 7 (2): 157-168.

[24] Oscar P, Paul R, Johnson-Roberson M, *et al.* Colquhoun Towards image-based marine habitat classification [C]// OCEANS 2008, IEEE, 2008: 1-7.

[25] Mary N A B, Dharma D. Coral reef image classification employing Improved LDP for feature extraction [J]. Journal of Visual Communication & Image Representation, 2017, 49 (nov.): 225-242.

[26] Qin Hongwei, Li Xiu, Liang Jian, *et al.* DeepFish: Accurate underwater live fish recognition with a deep architecture [J]. Neurocomputing, 2016, 187: 49-58

[27] Wei Guanqun, Wei Zhiqiang, Huang Lei, *et al.* Robust Underwater Fish Classification Based on Data Augmentation by Adding Noises in Random Local Regions. [C]// Pacific Rim Conference on Multimedia 2018: 509-518.

[28] Ramprasaath R. Selvaraju, Michael C, *et al.* Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization [C]// International Conference on Computer Vision, 2017: 1-24.

[29] Gao Yang, Beijbom O, Zhang Ning, *et al.* Compact bilinear pooling [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 317–326.

[30] Ammar M, Mohammed B, An Senjian, *et al.* ResFeats: Residual network based features for underwater image classification [J]. Image and Vision Computing, 2020. 1: Article ID 103811.

[31] 张俊龙，曾国荪，覃如符. 基于深度学习的海底观测视频中鱼类的识别方法 [J]. 计算机应用, 2019, 39 (2) 72-77. (Zhang Junlong, Zeng Guosun, Qin Rufu. Fish reconition method for submarine observation video based on deep learning [J]. Journal of Computer Applications, 2019, 39 (2) 72-77.

[32] Zhang Yu, Wei Xiushen, Wu Jianxin, *et al.* Weakly supervised fine-grained categorization with part-based image representation [J]. IEEE Transactions on Image Processing, 2016, 25 (4): 1713−1725.

[33] Perronnin F, Dance C. Fisher kernels on visual vocabularies for image categorization [C]// Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA: IEEE, 2007. 1−8.

[34] Simon M, Rodner E. Neural activation constellations: unsupervised part model discovery with convolutional networks [C]// Proceedings of the 15th IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1143−1151.